

compliance for these days exceeded 90%. Evidence for back and forward filling of paper diary cards was observed. For the compliance-enhanced electronic diary, the actual compliance rate was 93%.

**CONCLUSIONS:** Data from paper-based diaries are of questionable validity, given that many of their entries are not completed as required by the protocol. Science-based electronic diaries can produce high rates of patient compliance in the field. Improved methods for data collection should encourage researchers in the pharmaceutical industry to aggressively evaluate electronic PRO (ePRO) data to help differentiate their products.

## DS2

### POWER CALCULATIONS FOR WIDELY USED PATIENT-REPORTED OUTCOMES (PRO) MEASURES IN WOMEN'S HEALTH TRIALS

Abetz L<sup>1</sup>, Brandman J<sup>2</sup>, Plante M<sup>3</sup>

<sup>1</sup>Mapi Values, Cheshire, UK; <sup>2</sup>Pfizer Pharmaceutical Group, New York, NY, USA; <sup>3</sup>Pfizer Inc, Ann Arbor, MI, USA

**OBJECTIVE:** Increasingly, federal authorities are requesting power calculations for secondary endpoints in clinical trials, including patient reported outcomes (PRO). However, most PRO measures do not provide power calculations in their manuals; if provided, they are often based on mixed samples of males and females. It is well documented that female and male PRO scores often differ. Thus when designing women's health trials, it may be worthwhile to conduct power calculations using women's PRO scores and standard deviations. This study presents the power and sample-size calculations for a variety of questionnaires used in women's health studies.

**METHODS:** The Menopause Quality of Life questionnaire (MENQOL), Women's Health Questionnaire (WHQ), Psychological General Well-Being Index (PGWB), and Short Form 36 and 12 (SF-36/SF-12) were assessed. Published information on scores and standard deviations in female populations were used to determine sample sizes needed to detect differences between two experimental groups, post-intervention.

**RESULTS:** Results varied by questionnaire, due in part to varying score ranges across questionnaires. For example, to achieve 90% power with a ten-point difference the following sample sizes per treatment arm were required: 158 women when using the MENQOL vasomotor score (range: 0–100); 47 women when using the WHQ total score (range: 0–102); 70 women when using the PGWB total score (range: 22–132); 24 and 21 women when using the SF-36 and SF-12 Physical Component Summary (no floor/ceiling).

**CONCLUSION:** When calculating sample sizes, it is necessary to keep in mind the questionnaire's possible score range in order to ensure that the power calculation is based on a clinically meaningful difference between treatment groups. These results may be used to help calculate sample sizes needed to achieve sufficient power to detect statistically significant differences in women's health trials for these widely used measures.

## DS3

### DESIGN AND ANALYSIS OF UNIT COST ESTIMATION STUDIES: HOW MANY TYPES OF HOSPITALIZATION? HOW MANY COUNTRIES?

Glick HA<sup>1</sup>, Orzol SM<sup>1</sup>, Tooley JF<sup>2</sup>, Polsky D<sup>1</sup>, Mauskopf J<sup>3</sup>

<sup>1</sup>University of Pennsylvania, Philadelphia, PA, USA; <sup>2</sup>Pharmacia, Skokie, IL, USA; <sup>3</sup>Research Triangle Institute, Research Triangle Park, NC, USA

**OBJECTIVES:** The availability of reliable methods for imputing hospital costs (e.g., relative values from the U.S. diagnosis-related group (DRG) payment system, cost weights and average length of stay (LOS) estimates from the Australian refined DRG system, or average LOS from the French Patient Homogeneous Groups), allows one to address two design questions related to the collection of hospital unit cost data for use in multinational clinical trials with economic endpoints: 1) In individual countries, for how many types of hospitalization should estimates be obtained? 2) For how many countries should they be obtained?

**METHODS:** We addressed these questions by assuming that unit cost estimates for 47 types of hospitalization collected in four western European countries represented the universe of hospitalizations. This assumption provided a population estimate against which we could measure the error associated with imputations in samples drawn from the population. To answer the first question, we: 1) randomly sampled subsets of hospitalization types from this population and used them to develop imputation regressions; 2) used the results of the regressions to impute costs from the remaining hospitalization types, and 3) estimated measures of imputation error within each sample. To answer the second question, we performed a similar analysis, but instead sampled countries.

**RESULTS:** We found that the imputation error decreased as the number of types of hospitalization and countries sampled increased, but that the rate of reduction in error shrank. We also found that error was minimized by obtaining estimates for fewer types of hospitalization from more countries than the reverse.

**CONCLUSION:** The availability of reliable methods for imputing hospitalization costs allows one to economize on data collection. Our experiment suggested that collecting a small number of estimates (in our data, approximately 25) in as many countries as is feasible minimizes imputation error.

## SESSION III

### DIABETES

## DB1

### ESTIMATING HEALTH-CARE COSTS ASSOCIATED WITH DIABETES-RELATED COMPLICATIONS USING DATA FROM THE UNITED KINGDOM PROSPECTIVE DIABETES STUDY (UKPDS)

Gray A, Clarke P

University of Oxford, Oxford, UK